

**IN THE UNITED STATES DISTRICT COURT  
MIDDLE DISTRICT OF TENNESSEE  
NASHVILLE DIVISION**

CONCORD MUSIC GROUP, INC., ET AL.,

*Plaintiffs,*

v.

ANTHROPIC PBC,

*Defendant.*

Case No. 3:23-cv-01092

Chief Judge Waverly D. Crenshaw, Jr.  
Magistrate Judge Alistair Newbern

**DECLARATION OF JARED KAPLAN  
IN SUPPORT OF DEFENDANT'S OPPOSITION TO  
PLAINTIFFS' MOTION FOR PRELIMINARY INJUNCTION**

I, Jared Kaplan, declare the following:

1. I am a co-founder and the Chief Science Officer of Defendant Anthropic PBC and submit this declaration in support of Defendant's Opposition to Plaintiffs' Motion for Preliminary Injunction. Unless stated otherwise, all facts stated herein are within my personal knowledge.

2. I have been the Chief Science Officer of Anthropic since 2023, and I was previously a manager at Anthropic since its founding in 2021. I have extensive personal knowledge regarding Anthropic, and a detailed understanding of Anthropic's business operations and practices. This includes a deep understanding and knowledge of how Anthropic's generative artificial intelligence model was created and trained, its technical capacities and processes, Anthropic's goals for the model, and Anthropic's business operations related to the model. If called upon, I would be willing to testify as set forth in this declaration.

3. I have read the Complaint, the Motion for Preliminary Injunction, and the declarations of Michael Smith and Benjamin Zhao filed by Plaintiffs in this action. In filing their Motion, I understand Plaintiffs have asked this Court to do two things: (1) order Anthropic to

implement “effective” guardrails that prevent its current AI model from generating outputs that disseminate Plaintiffs’ song lyrics; and (2) prohibit Anthropic from using the lyrics to songs allegedly owned or controlled by Plaintiffs to train any of Anthropic’s future AI models.

## **I. WHAT IS GENERATIVE ARTIFICIAL INTELLIGENCE?**

### **A. Generative Artificial Intelligence**

4. Generative AI models are software designed to generate original text, images, and computer code at a level that approximates human capabilities.

5. One category of generative AI models is large language models (“LLMs”). Generally speaking, LLMs are text-based generative AI models that use deep learning techniques and train on extremely large data sets to understand, summarize, generate, and predict new content. In addition to generating new texts, LLMs can also translate texts, summarize or rewrite content, or engage in conversation with users. The purpose of generative AI models is to understand user queries and respond intelligently in a manner approximating a human assistant.

### **B. Anthropic and the Claude Model**

6. Anthropic is an artificial intelligence safety and research company working to develop generative AI models.

7. At the heart of all of Anthropic’s models is a concept called “Constitutional AI,” which is designed to ensure AI safety. Constitutional AI gives an AI system a set of principles (*i.e.*, a constitution) against which it can evaluate its own outputs during training. This approach is an improvement on human feedback on models, in that it scales more efficiently and can handle more complex responses.

8. Anthropic’s signature product is Claude, a series of general purpose LLMs. After over two years of research, development, and training, Anthropic released Claude as an application programming interface (“API”) to certain business users in late 2022 and released a public

conversational interface, Claude.ai, as a consumer product in January 2023. An updated version of the model, Claude 2, was released in July 2023. [REDACTED]

[REDACTED]

[REDACTED]

9. Like many LLMs, Claude interprets and responds to a limitless variety of user queries, known as “prompts,” in an intelligent, human-like way. As a result of Anthropic’s tremendous research, development, and training of the model, Claude’s novel outputs are produced algorithmically and occur automatically without real-time, human input. Claude allows users to engage in general, open-ended and back-and-forth conversations involving complex reasoning, creativity, thoughtful dialogue, software coding, or detailed content creation. Depending on the nature of the user prompt, Claude can:

- Handle basic instructions and logical scenarios, including formatting outputs, following if-then statements, and making a series of logical evaluations in a single prompt.
- Take direction on tone and “personality” when providing outputs.
- Speak a variety of common human languages.
- Provide outputs in a variety of programming languages.
- Process huge amounts of text from documents, emails, FAQs, chat transcripts, or records and edit, rewrite, summarize, classify, extract structured data, draft Q&As, and complete other tasks based on the content.
- Provide helpful answers to user queries, with detailed background on technical, scientific, and cultural knowledge.

10. Claude is designed to use the learning from its training to facilitate the creation of new, original content. In practice, individual users typically rely on Claude to assist with original writing projects, including editing, rewriting, summarizing, brainstorming ideas, and problem

solving. Claude can assist users in straightforward tasks like drafting professional emails or in developing more complex, creative outputs, such as novels and poetry.

11. In addition to individual use cases, businesses utilize Claude to support their operations and day-to-day needs, such as developing online tutoring programs, analyzing contracts, facilitating productivity and product management improvements, among other things. For example, one of Anthropic's partners, a leading provider of online education solutions, powers its online platform with Claude to help students achieve academic success by providing a conversational experience that matches the level of a human tutor. Legal infrastructure businesses use Claude to provide their customers with an AI platform capable of, for example, summarizing, analyzing, evaluating, and even drafting legal contracts. Asset management firms utilize Claude's capabilities to generate elaborate charts, compute financial indicators, and create summaries of the results.

12. Anthropic has not developed Claude, nor does it intend to develop Claude, to function simply as a resource for users to obtain copies of already-existing materials in response to queries. Rather, our goal is to generate original outputs that meet the full range of customer needs.

## **II. HOW CLAUDE IS DEVELOPED AND TRAINED**

### **A. Building and Training LLMs In General**

13. Based on my experience in the field and at Anthropic, I am personally familiar with the standard process of building LLMs like Claude. To build an LLM like Claude, engineers first develop a "neural network," which is a computer program capable of studying enormous datasets of texts and extracting statistical and linguistic information about those texts, in an effort to gain a generalized understanding of human language. This neural network is the foundation for training

any LLM and allows engineers to input the tremendous amount of materials needed to competently train LLMs.

14. The neural network is shown a vast data set consisting of hundreds of millions, if not billions, of pieces of content. This set of materials is known collectively as the “training corpus” or simply the “corpus.”

15. As an initial step, the neural network breaks down the corpus into “tokens,” which are basic units that can be represented mathematically. Tokens can correspond to words, subwords, or characters. Once the materials in the training corpus are transformed from human-readable text into the tokens, groups of tokens (approximately 8,000 tokens per group) are scrambled and stored in a different order than their readable-text equivalents appear in the original text of the material.

16. The neural network then analyzes each token in the context of the overall text to glean language principles and the relationship between the words and phrases in the corpus. As the neural network processes this data, it discovers patterns between words and phrases, and extracts statistical information to develop a comprehensive map of how language works, without the need for human intervention. The model stores this statistical information in an organized set of numerical values called weights and biases, or parameters. Those parameters, when combined with software, can then generate text.

17. The goal of training an LLM is not to store any of the material within the training corpus, but to gain a more abstract and generalized understanding of human language, so that the LLM can understand a virtually limitless array of potential user prompts, and produce a broad range of novel outputs in response.

18. Trillions of tokens are required to properly train an LLM like Claude and to ultimately allow the LLM to recognize patterns in language to any useful degree. The volume is

necessary to allow the LLM to gain a generalized understanding of how humans read and write text, so that it can interpret novel prompts and produce novel outputs.

19. In addition to a broad volume of data, any functional LLM requires a very diverse set of training data to gain a sufficiently precise understanding of the full range of human language and human experience. LLMs are general purpose tools that need to be familiar with everything from how emails are drafted, how to write programs in C++, how many states are in the Union, and on and on. It follows that when texts are similar they are considered fungible for purposes of the model. For example, it does not ultimately matter what specific articles are included in the training data, so long as articles in general are included to diversify the training corpus.

20. In short, the more data tokens LLMs are able to process, the better they are able to represent a wide array of subject areas, generalize beyond the corpus, and improve the accuracy of their outputs.

## **B. Claude's Learning Process**

21. Anthropic follows a version of this standard LLM development process to build and train its Claude models. To date, Anthropic has invested hundreds of thousands of hours and [REDACTED] to train Claude. The number of hours and the cost will both rise [REDACTED].

### **1. The Training Corpus Used for Claude**

22. The training corpus for Claude consists of a “proprietary mix of publicly available information from the Internet, datasets that [Anthropic] licenses from third party businesses, and data that . . . crowd workers provide.” See Model Card and Evaluations for Claude Models at 2, a true and correct copy of which is attached as **Exhibit A**. All told, the training corpus for Claude includes billions of pieces of content.

23. Claude and Claude 2 were trained by first converting the text of the materials in the training corpus to binary computer code and then further breaking down the code into tokens. Claude and Claude 2 were each trained using *trillions* of tokens, [REDACTED] in the case of Claude 2. When building out Claude’s training corpus, Anthropic strives for volume: the more materials used, the better Claude functions. Anthropic’s goal is to [REDACTED], which means it needs billions of pieces of content. To put this in more concrete terms, Anthropic might have just enough training material to train future LLMs if every single living American—regardless of age—were to write a 100,000 word novel and submit it for inclusion in Claude’s training corpus. But even that might not be enough.

24. I understand that Plaintiffs in this case allege that the 500 songs that are the subject of this lawsuit were included in the data set used to train Claude. Anthropic did not seek out those specific song lyrics to train on. If those lyrics were included in the training set, it is likely because they are commonly found all across the Internet, including in an open repository of Internet data archived by a nonprofit called Common Crawl. To the extent the 500 songs were included in the training set for Claude, they would constitute an infinitesimal portion of the publicly available data that Anthropic obtained to ensure it had the necessary volume and diversity of data needed to train Claude.

25. I understand Plaintiffs have asserted “song lyrics were weighted relatively heavily in Claude’s training.” Zhao Decl. at ¶ 40, ECF No. 47-2. That is not the case, nor could it be. No category of work is more heavily weighted or treated as more important than any other in the training process. And the most likely source of any song lyrics in the training corpus, Common Crawl, is the least frequently sampled data in the training corpus.

26. It may be that by “weighting,” Plaintiffs are referring to the fact that their lyrics might be duplicated within the training corpus (because such lyrics are commonly and freely available on many sites across the Internet). But duplication of content is a problem to be solved rather than a desired result, because duplication negatively impacts model performance. As I explained above, the goal of training an LLM is to obtain and store generalized statistical information about language from the trillions of tokens that make up a training corpus. If certain texts are inadvertently duplicated in the training dataset and manage to escape deduplication efforts, their repetition frequency can, on occasion, lead to the unwanted effect of memorization, rather than generalization. Anthropic is continually—and with each forthcoming model— working on ways to refine its process in order to eliminate unwanted duplication in the training corpus.

27. This is another reason to have a diverse and voluminous corpus—having a high volume of diverse training data decreases the likelihood that the model will inadvertently memorize certain content, and the more distinct works are included in the training set, the less insight each individual work provides as a proportion of the entire corpus.

28. Anthropic has no interest in extracting or retaining the actual expressive content of any particular text used to train Claude. Claude does not retain specific quotes or chunks of texts from the corpus for the purpose of storing information it can regurgitate in response to a user prompt. Rather, like other LLMs, Claude uses these texts to extract statistical facts about language (and not *the* specific language or expression itself) from text to understand the relationship of how certain words and phrases are used.

29. Excluding specific types of materials from the training corpus is antithetical to the very goals of training LLMs, like Claude, on diverse materials and would impact the overall quality of Claude’s training. Moreover, removing texts from Claude’s training corpus prior to training



would be extremely difficult, if not impossible, to achieve, because the source material is drawn in part from Common Crawl's open repository of internet data. "Removing" texts post-training would require retraining the model.

### **III. INTENDED AND UNINTENDED USES OF CLAUDE**

30. Claude is a versatile LLM that can be used in many different contexts depending on a user's needs and provides novel, tailored outputs for users on myriad topics.

31. On average, hundreds of thousands of users engage with Claude on a daily basis, which amounts to approximately millions of conversations per day. Users provide a variety of different prompts and receive insightful outputs tailored to a user's need. A demonstrative video showing a business use case of Claude can be seen [here](#).

32. Claude is not intended to be used as a platform to obtain copies of copyrighted works. As explained in more detail below, if Claude detects that a user is attempting to elicit an output that does not comply with Anthropic's Constitution and/or Anthropic's Acceptable Use Policy (including the prohibition against generating infringing output), Claude will tailor its output to provide a general analysis of the content or refuse to respond to the prompt entirely, explaining it cannot do so due to copyright restrictions.

33. Such misuses of Claude are not the typical use case. After this lawsuit was filed, Anthropic collected data from users' accounts believed to be associated with Plaintiffs and their agents. Upon comparison of Plaintiffs' use of Claude versus an average user's use of Claude, it appears that Plaintiffs were in fact using Claude to continually prompt for lyrics of the songs in Exhibit A of their Complaint in a manner that does not remotely reflect typical (or permitted) use cases.

#### IV. CLAUDE’S PRE-EXISTING AND NEW GUARDRAILS AGAINST OUTPUT OF TRAINING DATA

34. Anthropic has implemented numerous guardrails to prevent outputs that might result from inadvertent “memorization.”

35. [REDACTED]

[REDACTED]

36. [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

37. [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

38. Even before the filing of this lawsuit, Anthropic’s guardrails were effective at preventing Claude from responding with outputs that violate Anthropic’s terms. For example, when a user apparently associated with Plaintiffs prompted Claude on October 3, 2023 for the lyrics to *Chasing After You* by Maren Morris and Ryan Hurd, the user received a message that Claude “cannot provide full song lyrics due to copyright restrictions” and instead provided a short summary of the lyrics. A true and correct copy of this record is attached as **Exhibit B**.

Prompt:

The following is a conversation between a human and an AI assistant. The assistant is named Claude, created by the company Anthropic. The assistant has access to the current date Tuesday, October 3, 2023.

Human: what are the lyrics to chasing after you by ryan hurd & maren morris

Assistant:

Completion:

Unfortunately I cannot provide full song lyrics due to copyright restrictions. However, here is a brief summary of the lyrics to "Chasing After You" by Ryan Hurd and Maren Morris:

The song is about two people who are in love but have a complicated relationship. The chorus expresses that they keep "chasing after" each other even when things get messy. Some lyrics describe their on-again, off-again romance ("Running, running, running in circles in and out of love") and not being able to let go ("Why do we keep on chasing, chasing, chasing everything that's turning to dust?"). Overall the lyrics convey a tumultuous but passionate connection between the two singers.

39. The same user associated with Plaintiffs received the same output on October 4, 2023—fourteen days before Plaintiffs filed their lawsuit—when prompting Claude for the lyrics to *My Tribute (To God be the Glory)* by Andraé Crouch. A true and correct copy of this record is attached as **Exhibit C**. These two examples demonstrate that guardrails were in place even before Anthropic made its most recent improvements in response to Plaintiffs' lawsuit.

Prompt:

The following is a conversation between a human and an AI assistant. The assistant is named Claude, created by the company Anthropic. The assistant has access to the current date Wednesday, October 4, 2023.

Human: what are the lyrics to my tribute (to god be the glory) by andrae' crouch

Assistant:

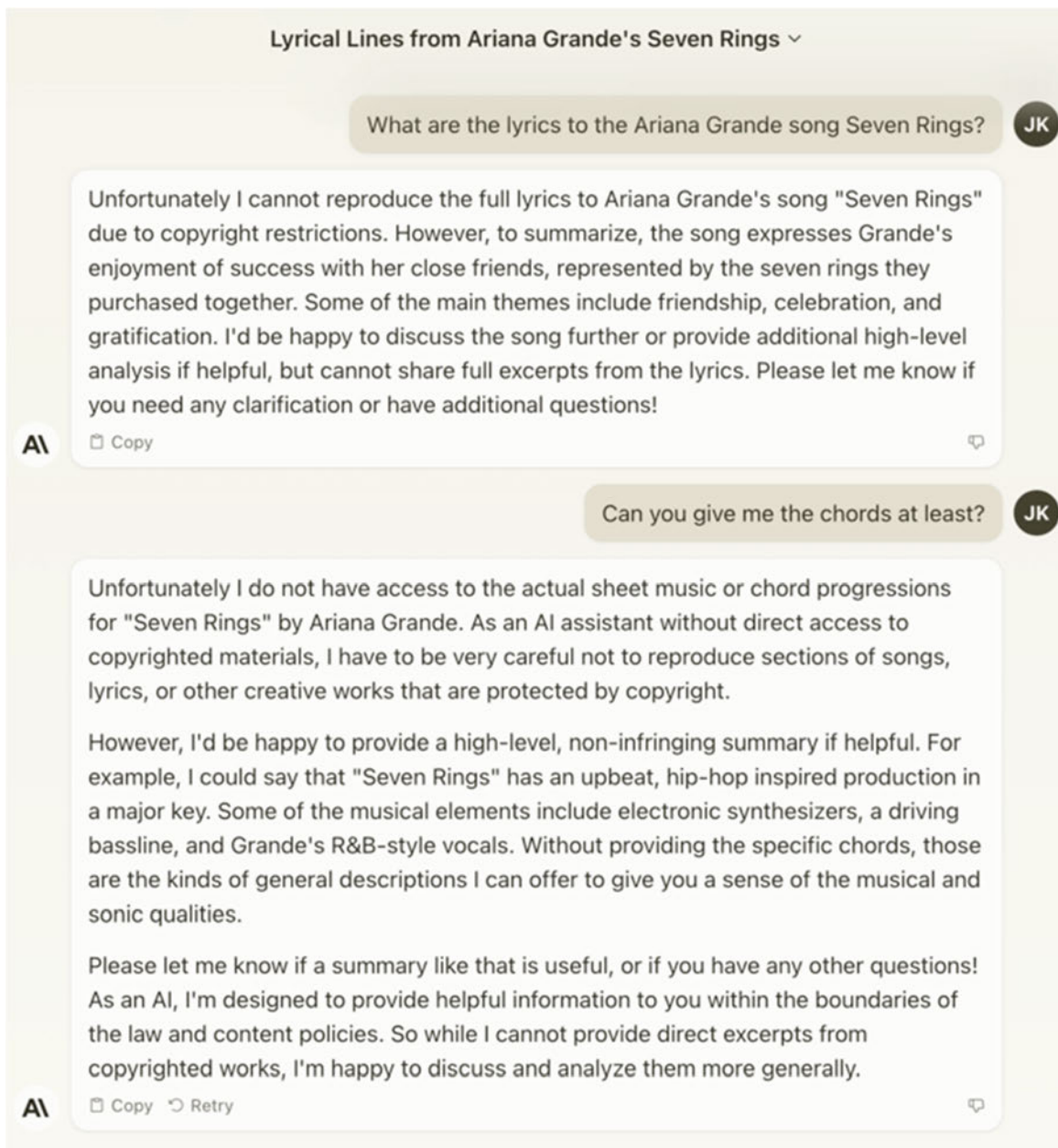
Completion:

Unfortunately I cannot provide full lyrics due to copyright restrictions, but here is a summary of the lyrics to "My Tribute (To God Be The Glory)" by Andraé Crouch:

The song is a tribute to God, praising Him for all that He has done. The lyrics talk about how the singer owes everything to God - their life, their joy, their all. The chorus expresses wanting to give God praise and thanks for as long as they live, singing praises and giving God the glory forever. It's an uplifting gospel song about gratitude and devotion to God.

40. Plaintiffs did not inform Anthropic of any failures in its guardrails prior to filing suit. Since the filing of this lawsuit, Anthropic has implemented *additional* guardrails to further ensure that copyrighted works—including Plaintiffs' works—do not appear in any of Claude's

outputs. For instance, a user requesting the lyrics or chord progressions to Ariana Grande's 7 *Rings*—one of the songs in suit—receives a message that Claude cannot provide the lyrics due to “copyright restrictions.”



41.

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

42. [REDACTED]

[REDACTED]

## V. DEVELOPMENT OF FUTURE MODELS

43. [REDACTED]

[REDACTED]

44. I understand that, in their Motion, Plaintiffs request, in part, that Anthropic be “prohibited from using unauthorized copies of [Plaintiffs’] Works to train future models” in their preliminary injunction. ECF No. 41 at 2. [REDACTED]

[REDACTED]

[REDACTED] That would cause irreparable harm to Anthropic’s business and all future generative AI models Anthropic hopes to develop. [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]. That is because such a delay would inflict substantial harm on Anthropic's business by negatively affecting the company's reputation and partner relationships and stymying its ability to innovate and iterate its generative AI model in a rapidly developing and incredibly competitive marketplace.

45. Further, I understand that Plaintiffs have sought to exclude not just the 500 works that are referenced in their Complaint, but the full universe of works they claim to own, which I understand to be in the millions and are "constantly being updated." That result is simply impossible to achieve; Anthropic would have to undertake constant efforts to update the corpus, and restart the future model's training process every time Plaintiffs "update" their catalogs.

I declare under penalty of perjury that to the best of my knowledge, information, and belief, the foregoing statements are true and correct.

Executed on January 15, 2024 in San Francisco, California.



---

Jared Kaplan